

TEXT-TO-SPEECH SYSTEM FOR TELANGANA STATE LANGUAGES

M. NAGAMANI¹, SYED ABDUL HAMEED², DHEERAJ KUMAR², V. S. VARUN²,
E. KUMAR GOUD², P. TAGORE FINNY², T. SAI SRUJANA² & M. SAIDI REDDY³

¹School of Computer and Information Sciences, University of Hyderabad,

Hyderabad, Telangana, India

²Students, Department of Computer Science, KG Reddy College of Engineering and Technology,

Hyderabad, Telangana, India

³Department of Computer Science, KG Reddy College of Engineering and Technology,

Hyderabad, Telangana, India

ABSTRACT

TTS Speech conversion system makes digital media comfortable by allowing miniature of devices with the help of speech interactive system. Through Text-to-speech can access to digital media with the help of speech modality. They can quarry by written terms and access information in their mother tongue as a form of text that in turn convert in speech by TTS. The present work explores the development of the two state languages of Telangana i.e., Telugu and Urdu in the context of speech synthesis and their application with the prosody manipulation to get naturalness in their language context. Also done a preliminary work using speech synthesis to implement language learning application of Telangana language. Hence, the name TTS System for Telangana state languages. The state language of Telangana Telugu and Urdu, teaching in primary school in the scope of this work. The limited domain application the implemented Urdu and Telugu speech synthesis and observed spoken term naturalness by human observation. There are many open source tools available for TTS but the most commonly used is festival as it is easy for users to modify the code and manipulate it.

The development of Urdu TTS through the transliteration in Telugu and use the same package for both the languages.

KEYWORDS: Dual-Lingual TTS, Urdu Transliteration, Speech synthesis, Telangana State Languages & Festival

Received: Aug 13, 2018; **Accepted:** Sep 03, 2018; **Published:** Nov 15, 2018; **Paper Id.:** IJCSSEITRDEC20181

1. INTRODUCTION

The word 'Synthesis' is defined by the Dictionary as 'the procedure of combining parts or elements so as to form a whole'. Speech synthesis generally refers to the artificial way of generating human speech for any devices. The devices that are used for synthesis are called as 'Speech Synthesizer', it may be either hardware based and software based. A Text-To-Speech synthesizer (TTS) is a computer-based program in which the system processes through the text and reads it aloud. The Text-to-Speech (TTS) synthesis procedure consists of two main phases. The first one is text analysis, where the input text is transcribed into a phonetic or some other linguistic representation, and the second one is the generation of speech waveforms, where the acoustic output is produced from this phonetic and prosodic information. These two phases are usually called as high- and low-level synthesis.

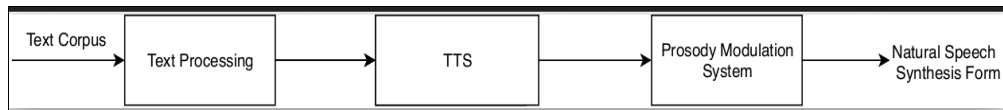


Figure 1: Simple Text-To-Speech Synthesis Procedure

The Text-To-Speech synthesis (TTS) is implemented entirely in software and only standard audio capability is required. At present, it contains several components, each of which handles a different task. For example, the text analysis capabilities of the system detect the ends of sentences, perform some rudimentary syntactic analysis, expend digit sequence into words, and disambiguate and expand abbreviation into normally spelled words which can then be analyzed by the dictionary-based pronunciation module.[1] In many applications like electronic mail messages, and generating spoken prompts in voice response system, there is a lots of demands of technology, which produces good and acceptable speech. The performance and quality of the Speech Synthesizer can be measured based on its naturalness and ability to be understood by its listener.

There are many speech synthesis techniques but most widely used are unit selection and Hidden Markov Model (HMM).

1.1 Unit Selection

It is the most widely used speech synthesis technique in which the text is dived into parts i.e., individual syllables, words, phrases, phones and diphones. During synthesis, the synthesizers utilize the information related to units, and pick the most appropriate unit based on the target cost and the concatenation cost. On the basis of target cost, best match units in the database are identified, whereas the joining cost choose the units that can be concatenated smoothly. The best optimal selected units are concatenated and speech is synthesized.

1.2 Hidden Markov Model

The HMM based speech synthesis framework performs simultaneous modelling of pitch and spectrum taking into account the dynamics of both quantities as well. Spectral representation utilizes Mel-based Cepstral coefficients while prosody is represented as logF0. Multi Space probability Distribution (MSD) modelling is performed to alleviate the problem of non continuous pitch values in unvoiced regions. Moreover, context clustering is performed using decision trees so as to fully exploit the contextual information in lexical and syntactic level

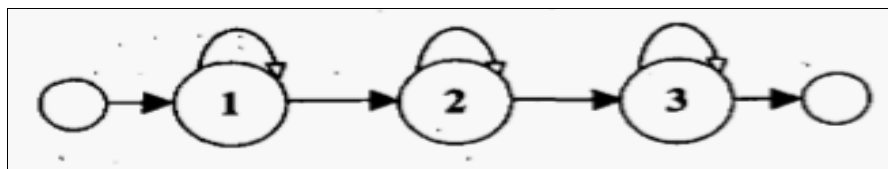


Figure 2: Hidden Markov Model (Tri-State)^[6]

2. FESTIVAL

The festival is a tool which offers a general framework for building speech synthesis systems (TTS) as well as including examples of various modules. As a whole it offers full text to speech through a number APIs: from shell level, though a Scheme command interpreter, as a C++ library, and an Emacs interface. The main usage of the Festival is to convert the text file or any text input into voice (Speech). When you pass a text file to the Festival it converts the contents of a text file into voice. For example, if I want to read a letter (mail) which is residing in a text file (say letter.txt).

It can let festival read it out loud for me as follows: **festival >(tts “filename.text”, nil)** or **festival>(SayText “telugu bhasha”)**. The advantages of this tool is, it is available for free under open source license, and In festival the voice quality and pronunciation are good and understandable.

To install Festival in our computer or laptops, open the Terminal (Ctrl+Alt+T), then type the command as follows:

\$sudo apt-get install festival

After installation of festival tool, type \$festival in Terminal to check whether it is installed or not. If it was installed then it will produce the window as shown below:

```
dheeraj@dheeraj-Aspire-E5-576:~$ cd Desktop/project_HCU/speech/
dheeraj@dheeraj-Aspire-E5-576:~/Desktop/project_HCU/speech$ festival
Festival Speech Synthesis System 2.4:release December 2014
Copyright (C) University of Edinburgh, 1996-2010. All rights reserved.
clunits: Copyright (C) University of Edinburgh and CMU 1997-2010
clusterngen_engine: Copyright (C) Carnegie Mellon University 2005-2014
hts_engine:
The HMM-Based Speech Synthesis Engine "hts_engine API"
hts_engine API version 1.07 (http://hts-engine.sourceforge.net/)
Copyright (C) The HMM-Based Speech Synthesis Engine "hts_engine API"
Version 1.07 (http://hts-engine.sourceforge.net/)
Copyright (C) 2001-2012 Nagoya Institute of Technology
2001-2008 Tokyo Institute of Technology
All rights reserved.
All rights reserved.
For details type `(festival_warranty)'
festival>
```

Figure 3: Screenshot of Festival TTS Prompt to Convert Given Text into Speech Utterance

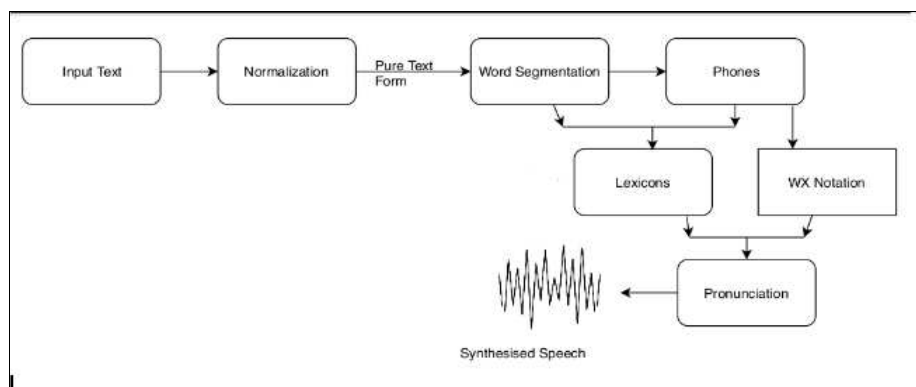


Figure 4: Block Diagram of Speech Generation Using Festival Speech Synthesis Tool

3. THE LANGUAGES OF TELANGANA

The state languages of the Telangana are Telugu and Urdu, as most of the population speaks both the language and these both languages have their own heritage and respect among itself. So as to preserve our state languages we are here with the idea of developing a TTS system for both the languages Telugu and Urdu. We are using the tool called festival, which is an open-source and easy to use for TTS system. As most of the primary schools has the state language as a mode of teaching so the students can directly learn these languages, and the problem of dual-language implementation will be eliminated.

4. PROCESS OF EXECUTING THE TTS SYSTEM FOR NATIVE LANGUAGES OF TELANGANA

4.1 Telugu Implementation

The language package tool is extracted and can be installed through command prompt using the following command. The language name is given in the command as 'te' and 'nsk' is the speaker information. Male voice features are used to produce the speech signal from the given input text by using the above block diagram shown in Figure 3:

4.1.1 \$sudo apt-get Install Festival-te-nsk

Once the package for the language used, then the next step to call the function from the command prompt to use the speaker voice. Command at the prompt is written in a scheme language with voice details shown below. The command will change the language from English US (default) to Telugu:

```
$festival>(voice_telugu_NSK_diphone)
```

where 'voice_telugu' is the telugu voice (language) and the 'diphone' are the phonemes of telugu language that is used to build the package. But the problem with this package is, it's not synthesized (no naturalness) and the pronunciation cannot be understand. To bring the naturalness in this package we have to control its speed (duration) of utterance as follows (For the speed of utterances):

```
$festival>(Parameter.set 'Duration_Stretch' <numeric value>)
```

here the speed of the utterance is indirectly propositional to numeric value (lower the value higher the speed rate). To make system repeat same utterances (same sentences) we have to store the utterances by 'set' and the 'no' indicates how many times you want it to repeat. For this we have type a command (This command will store utterance in utt and also produce the voice) in Terminal as follows:

```
$festival>(set! utt<no> (SayText "welugu bhasha"))
```

And to repeat the same utterance for multiple times we can use this command, where 'no' is number of repetition you want:

```
$festival>(utt.play utt<no>)
```

4.2 Urdu Implementation

The Urdu language package is not available in the festival so we are here with implementing this language using its transliteration in Telugu so that by using the same package both languages can be implemented. The Urdu language can be written in Telugu words (transliteration) and can be further modified to generate the natural synthesized speech.

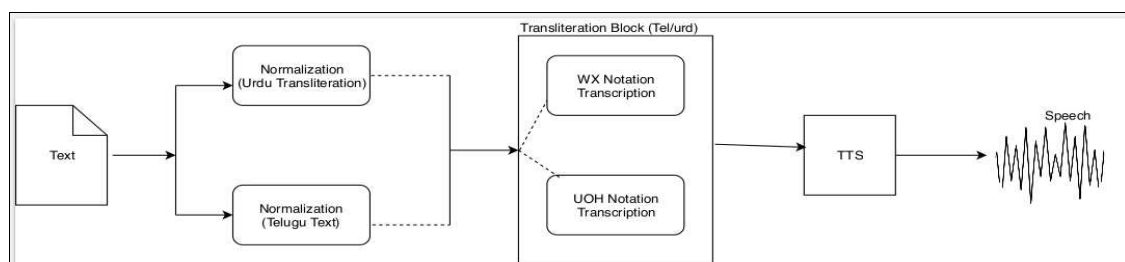


Figure 5: Implementation of Bi-Lingual System using the Same Package

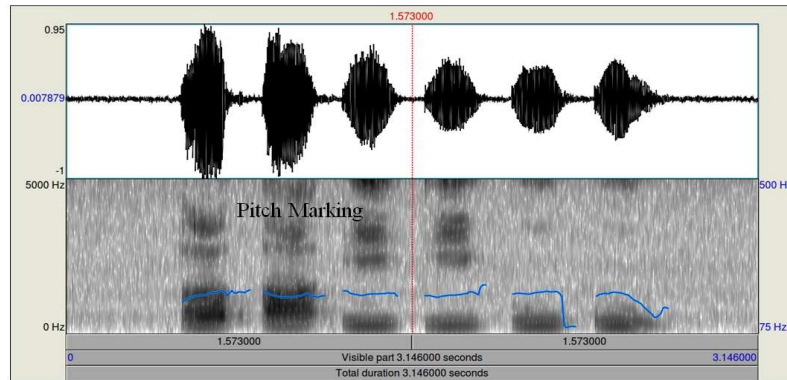


Figure 6: Speech Signals of Below Given Urdu Text

خوش آمدید = kushu AmxIx (Kush amdeed)

میں ٹھیک ہوں، شکر یہ = mai wIk hu ShukrIya (Mai theek hu shulriya)

Table 1: Speech Signals of Telugu Phoneme Sounds with Pitch Marking in Blue Line for Below Given Text with Male VOICE

Telugu Text	English Transliteration	UOH Notation
అ	a	AX
ఆ	aa	AA
ఇ	e	IX
ఈ	ee	IY
ఊ	ou	UH
ఋ	ouu	UA

5. CONCLUSIONS

The TTS system using a festival schema language and making system to generate the voice in a more natural format. The present work is by using a festival tool to manipulate speech prosody for Telangana State languages and the future scope of work is incorporating user-defined phonetic representation in language to bring naturalness as per nativeness, collecting a large text corpus of Urdu and using a translation program in Telugu to build a text corpus and speech corpus to extract speaker voices of Telugu and Urdu naturalness. The development project will help E-learning in primary level education and enhances the present multi-lingual language teaching in Telangana State school syllabus.

REFERENCES

1. *HMM-Based Speech Synthesis for the Greek Language*-- <https://www.researchgate.net/publication/221152134>
2. *BHAASHIKA: TELUGU TTS SYSTEM*-- Dr. K. V. N Sunitha et. al. / *International Journal of Engineering Science and Technology* Vol. 2(11), 2010, 6269-6277
3. *The HMM-based Speech Synthesis System (HTS) Version 2.0--6 th ISCA Workshop on Speech Synthesis, Bonn, Germany, August 22-24, 2007*

4. <http://www.festvox.org/festival/>
5. *Comparison of Urdu Text to Speech Synthesis using Unit Selection and HMM based Techniques—2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Technique (O-OCOSDA) 26-28 October 2016, Bali, Indonesia*
6. *A Human Machine Speaker Dependent Speech Interactive System --INDIAN INSTITUTE OF TECHNOLOGY. KHARAGPUR 721302, DECEMBER 20-22,2004*
7. *Improving Reading and writing skills with Intelligent Tutor for Telugu Language Learning INTELL M. Nagamani, M. Narendra prasad and P. N Girija.*
8. *Gavhane, S. P., & Shelake, V. M. Intrusion Detection System using Optimal C4. 5 Algorithm.*
9. *M. Nagamani, P. N. Girija., “Investigations of features for audio Information Retrieval ”, Journal of Acoustic Society of India, Vol.32 No. 1-2, 2004.(Introduction chapter)*
10. *G. Sreenu, P. N. Girija, M. Nagamani, M. Narendra Prasad, “Telugu Speech Interface Machine for University Information System”, Proc. Of IEEE Advanced Computing and Communications (ADCOM), Ahmadabad, December 2004. (scopus i.e IEEE explorer*